Research Data Curation Program: Building Buzz at UC San Diego

Reid Otsuji

19th Conference of ASLI, 2016

New Orleans, LA

The Library
UC SAN DIEGO

What is all the buzz about?

The Library
UC SAN DIEGO

Part 1
Who we are. . .

What are we doing?

Why are we doing it?

Who are we doing it for?

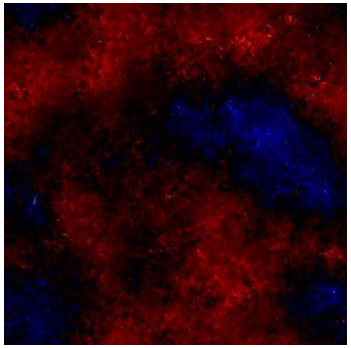
Part 2
The Library Research Data Digital Collections

Buzzing Ahead. . .

Image: http://library.ucsd.edu/dc/collection/bb5291666j

Research Data Curation Program: The Brief History

- 2 year pilot program beginning in 2011.
- Learn how researchers, information technologists, and librarians work together with data.
- In 2013, the library reorganized and the Research Data Curation Program was a new addition to the program structure.



The Library
UC SAN DIEGO

Who we are. . .now

Old team:

UCSD Libraries Team

- Ardys Kozbial, Director, Research Data Curation Program
- David Minor, Head of Curation and Preservation Services, San Diego Supercomputer Center
- Mary Linn Bergstrom, Curation Analyst and Faculty Liaison
- Sue McGuinness, Faculty Liaison
- Brad Westbrook, Metadata Consultant
- Declan Fleming, Technology Consultant

http://libraries.ucsd.edu/services/data-curation/contact-us.html

New team:

Research Data Curation Program Team

Our team is here to provide you with a variety of services and support.

Contact the Research Data Curation Program with any questions or comments. Feel free to email any of us individually if we can be of service.

David Minor is the Program Director for Research Data Curation. He oversees general direction for the group, and leads strategic partnerships on campus and within the UC. He's also Director of the Chronopolis Program, a long-term digital preservation network. (ORCID, LinkedIn, ResearchGate)

Ho Jung Yoo is a Technical Analyst for Research Data Curation. She works directly with researchers seeking to manage and share their data. She brings to the program years of research experience in population ecology. (ORCID, ResearchGate)

Ryan Johnson is a Metadata Specialist for Research Data Curation. He helps researchers describe their data and makes sure it's discoverable in the digital repository. He has experience in managing metadata for digital humanities and social science data.

Juliane Schneider is a Research Data Strategist for Research Data Curation. She works with the processes involved in the data lifecycle, including the areas of discoverability, ontologies, data sharing and the federal requirements for data sharing and data management plans. Her experience includes development and use of ontologies in discoverability, development of tools for text mining, and data analysis/transformation.

Mary Linn Bergstrom is a Liaison Librarian. She manages the EZID service, funded by the Library, whereby campus researchers can set up accounts and assign digital object identifiers (DOIs) to their research datasets. She also participates in developing and teaching workshops on the concepts of research data management and curation. (ORCID)

Reid Otsuji is a Data Curation Specialist and Faculty Liaison Librarian for Research Data Curation. He helps campus researchers with creating data management plans using the DMPTool and designs and teaches curriculum for data curation-related training and workshops. (ORCID)

Sibyl Schaefer is the Digital Preservation Analyst for Research Data Curation. She is Program Manager for Chronopolis and helps define long-term digital preservation solutions for the UC San Diego campus, also continuing our commitment to the Digital Preservation Network (DPN) and the National Digital Stewardship Alliance (NDSA). (LinkedIn)

Tim Dennis is the Library Data Librarian. He provides consultations on research involving data, including finding and recommending data sources and advising on technical data issues, such as file format conversion, web scraping, and statistical software use. He teaches workshops on various data and programming topics, including software use, data manipulation, management, and literacy. He also gives in-class presentations to introduce students to data sources and statistical software. (LinkedIn, ORCID)

What are we doing?

Building strategic partnerships.

Data Science Student Society

School of GLOBAL POLICY & STRATEGY

INTEGRATED DIGITAL INFRASTRUCTURE (IDI)

QUALCOMM INSTITUTE

Calit2

SCRIPPS INSTITUTION OF OCEANOGRAPHY
UC San Diego

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

What are we doing?

Promotion and Outreach.

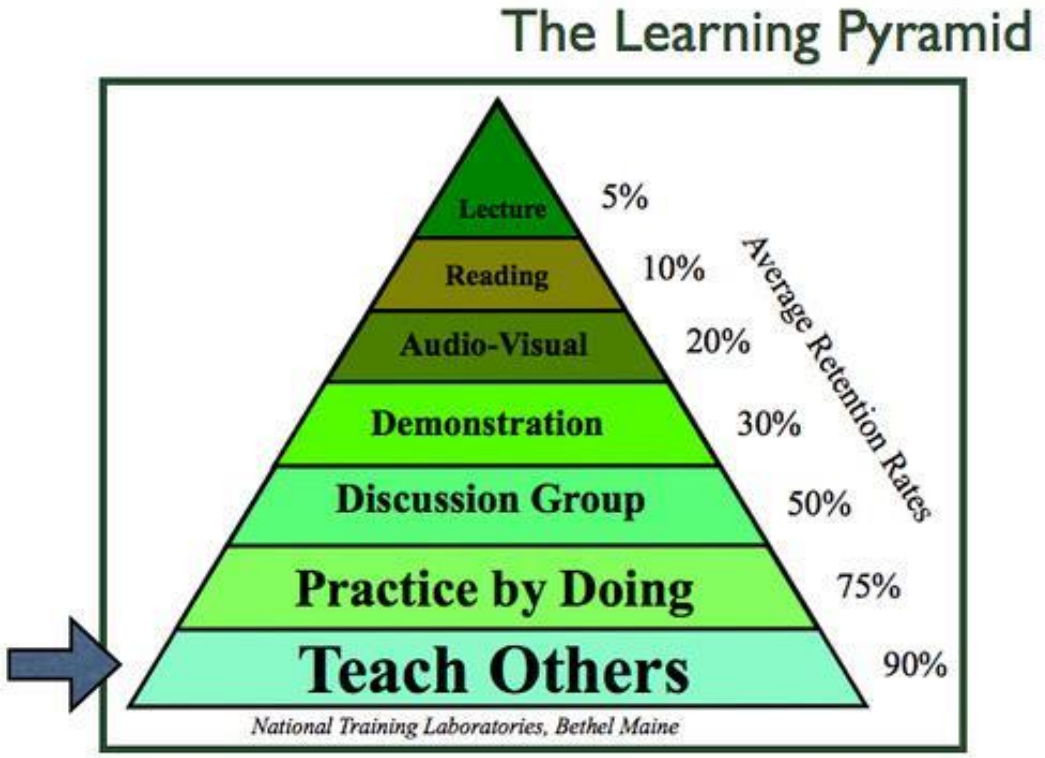
- Hosting luncheon forums.
- Website redesign.
- Giving away promotional swag.
- Tabling at library sponsored events, new faculty, graduate, and post-doctoral orientations.
- Posting flyers on bulletin boards around campus.
- Networking at local research related events.
- Working with our liaisons librarians.

The Library
UC SAN DIEGO

What are we doing?

Developing training.

- Build data management training plan.
- Library Learning Services Program – developed by trained instructional designers.
- Instructional designers developed a comprehensive Learning Support Plan (LSP) for us to use as a resource for training development.



The Library
UC SAN DIEGO

Improved Training, From this . . .

Initial training plan started with one-shot workshops for librarians and faculty researchers.

Before Learning Support Plan (LSP)	Audiences	Training Support Solutions	
		Learning Resources	People Resources
	Faculty Researchers	Workshops	Liaison Support
Data Management	●	●	●

DMPTool
Build your Data Management Plan

The Library
UC SAN DIEGO

To this . . .

Instructional designers created a Learning Support Plan matrix that outlined where we should focus our resources.

Use RDCP Services and Resources to Apply Best Practices of Data Management		Audiences				Learning Support Plan						
						Reference Resources			Learning Resources		People Resources	
		Librarians	Faculty/ Researchers	Graduate Students	Undergraduates	FAQ	Job Aids (Steps)	Policies	Online Demos	Workshops	Consultations	Liaison Support
Tasks and Processes	1. Navigate website	●	●	●	●	●			●	●	●	●
	2. Establish DMP tool account		●	●			●		●		●	
	3. Create EZID		●	●			●		●		●	
	4. Create naming convention		●	●		●	●				●	
	5. Create file structure		●	●		●	●				●	
	6. Create metadata		●	●			●			●	●	
	7. Create DOIs/ORCID		●	●			●		●		●	
	8. Deposit data		●	●			●		●		●	
	9. Use storage depository (DAMS)		●	●		●	●		●	●	●	
	10. Use long term storage (Chronopolis)		●	●		●	●		●	●		
Concepts	1. Understand the data management lifecycle	●	●	●	●				●	●	●	●
	2. Create a Data Management Plan	●	●	●				●	●	●	●	●
	3. Understand preservation and the value of easy access sharing	●	●	●	●				●	●	●	●
	4. Establish workflows		●	●					●	●	●	
	5. Understand funder mandates	●	●	●		●		●		●	●	●
	6. Understand storage options	●	●	●					●	●	●	

To This . . . Introducing Data Analysis Tools!

https://software-carpentry.org/

DATA CARPENTRY

MAKING DATA SCIENCE MORE EFFICIENT

http://www.datacarpentry.org/

Exploring the need to teaching popular tools used in data science for analyzing and cleaning data.

GitHub

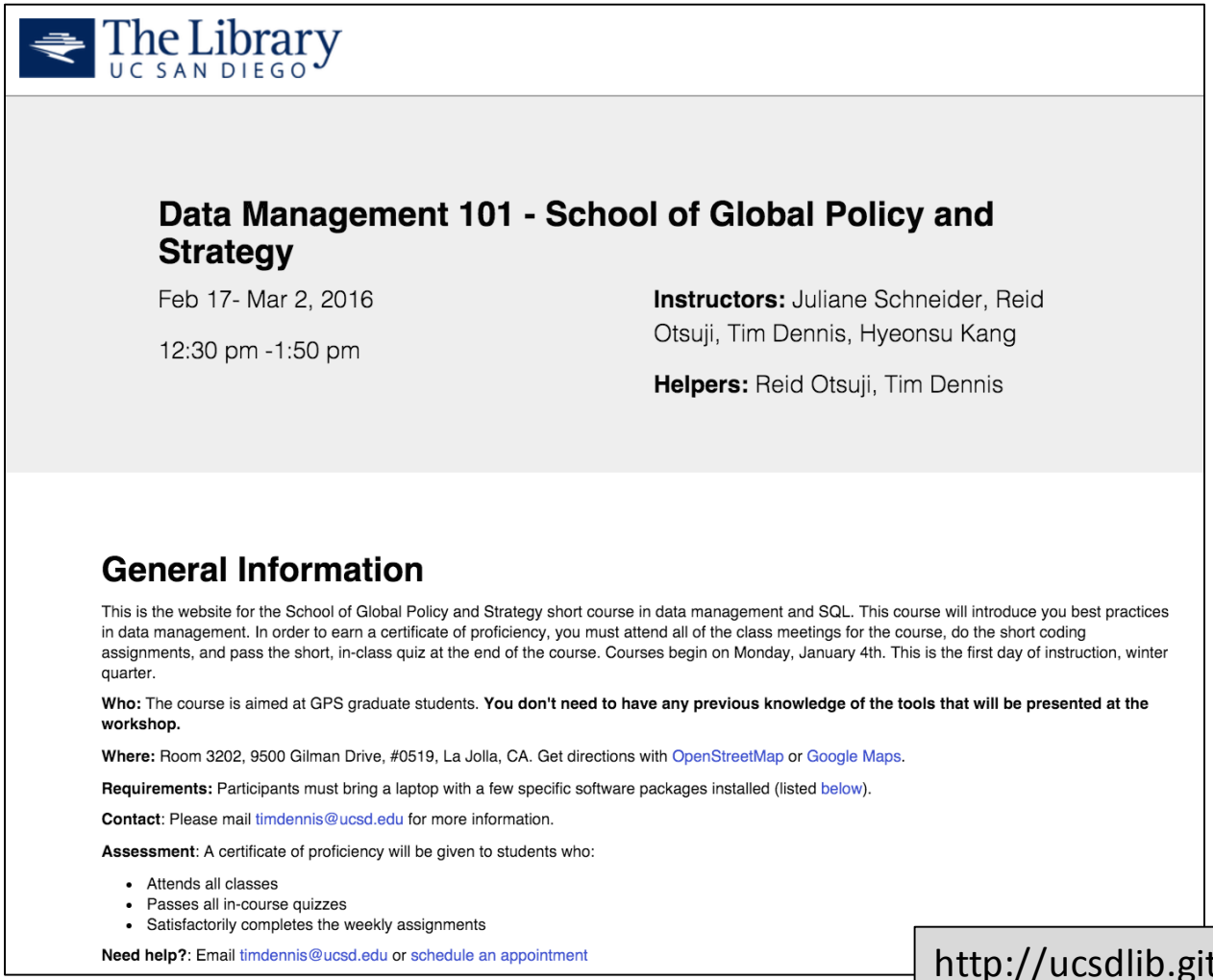
python™

R Studio

The Library
UC SAN DIEGO

Which Lead To This…

The Library UC SAN DIEGO



The Library UC SAN DIEGO

Data Management 101 - School of Global Policy and Strategy

Feb 17- Mar 2, 2016

12:30 pm -1:50 pm

Instructors: Juliane Schneider, Reid Otsuji, Tim Dennis, Hyeonsu Kang

Helpers: Reid Otsuji, Tim Dennis

General Information

This is the website for the School of Global Policy and Strategy short course in data management and SQL. This course will introduce you best practices in data management. In order to earn a certificate of proficiency, you must attend all of the class meetings for the course, do the short coding assignments, and pass the short, in-class quiz at the end of the course. Courses begin on Monday, January 4th. This is the first day of instruction, winter quarter.

Who: The course is aimed at GPS graduate students. **You don't need to have any previous knowledge of the tools that will be presented at the workshop.**

Where: Room 3202, 9500 Gilman Drive, #0519, La Jolla, CA. Get directions with OpenStreetMap or Google Maps.

Requirements: Participants must bring a laptop with a few specific software packages installed (listed below).

Contact: Please mail timdennis@ucsd.edu for more information.

Assessment: A certificate of proficiency will be given to students who:

- Attends all classes
- Passes all in-course quizzes
- Satisfactorily completes the weekly assignments

Need help?: Email timdennis@ucsd.edu or schedule an appointment

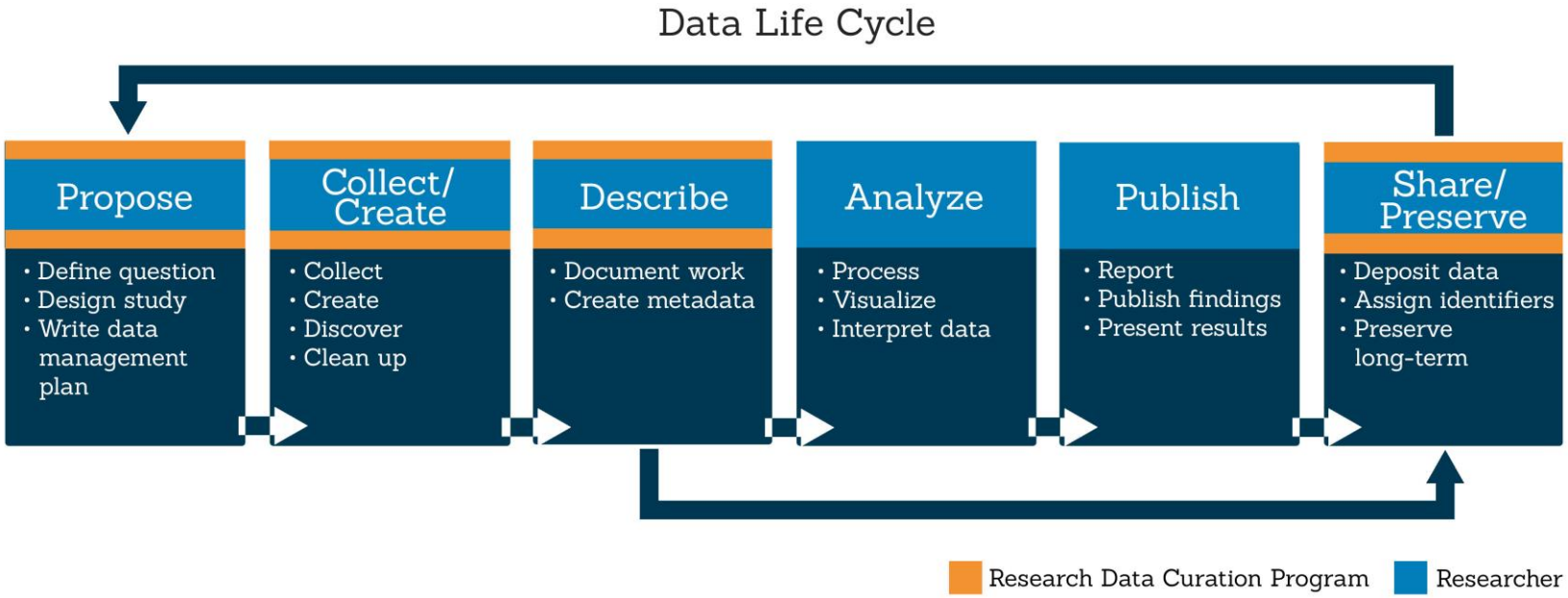
School of GLOBAL POLICY & STRATEGY

http://ucsdlib.github.io/win2016-gps-intro-R/

http://ucsdlib.github.io/win2016-gps-dm101/

Why are we doing it?

- Management of data throughout its lifecycle …

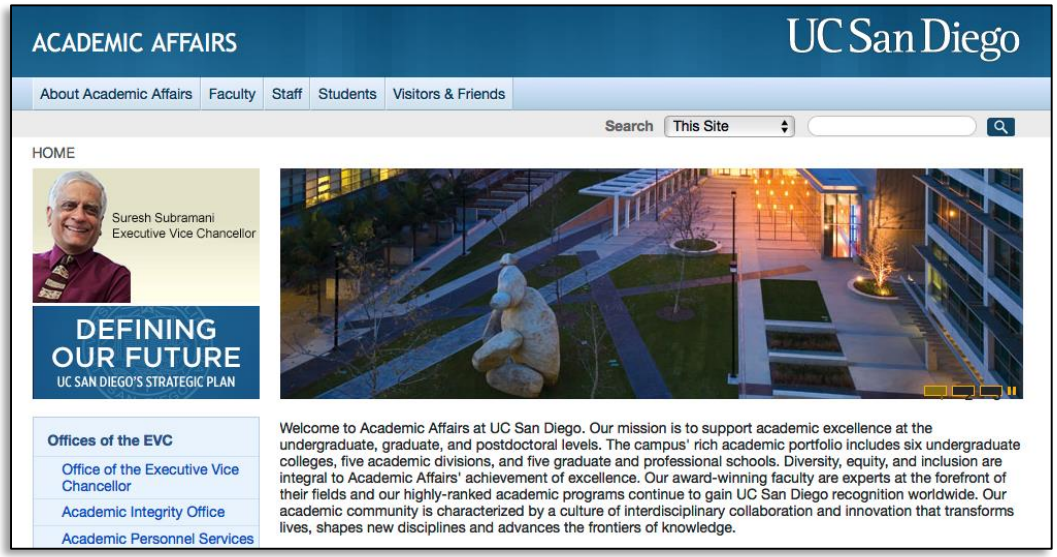


… to ensure it is reliably retrievable for future research purposes or reuse.

The Library
UC SAN DIEGO

Who are we doing it for?

The Library
UC SAN DIEGO

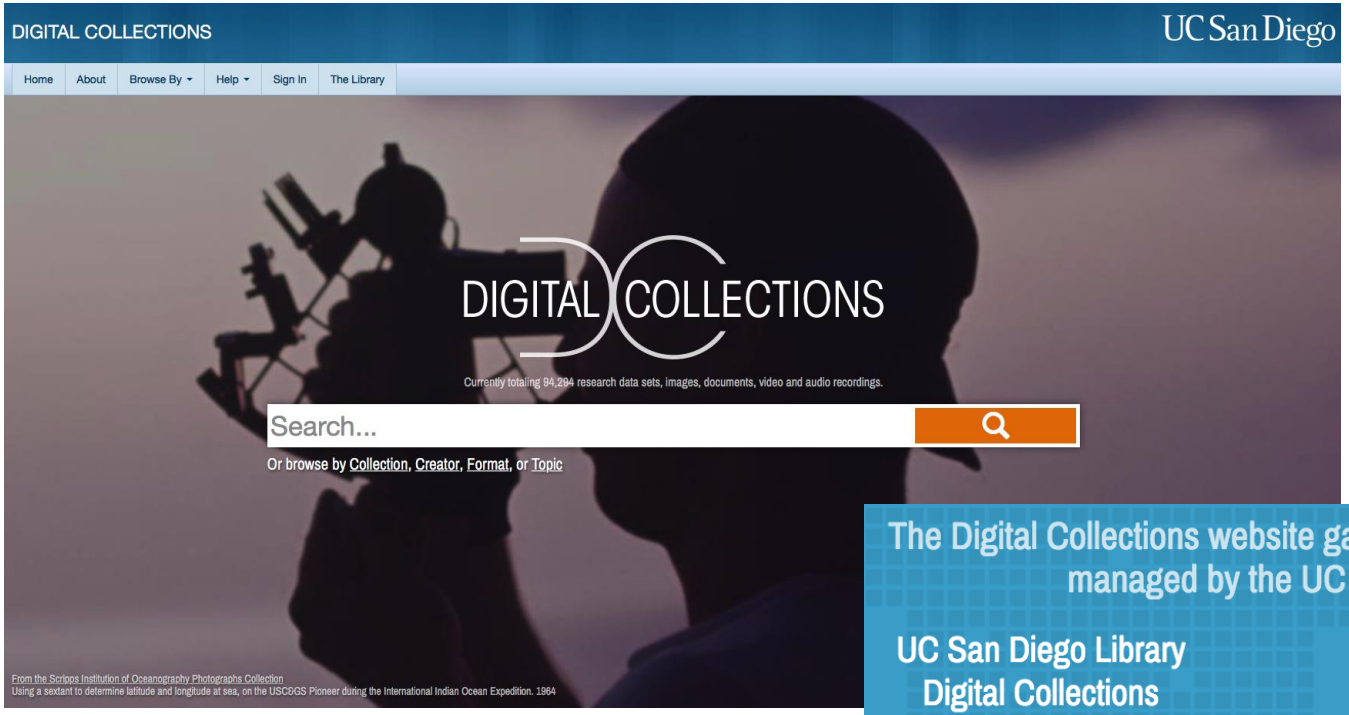


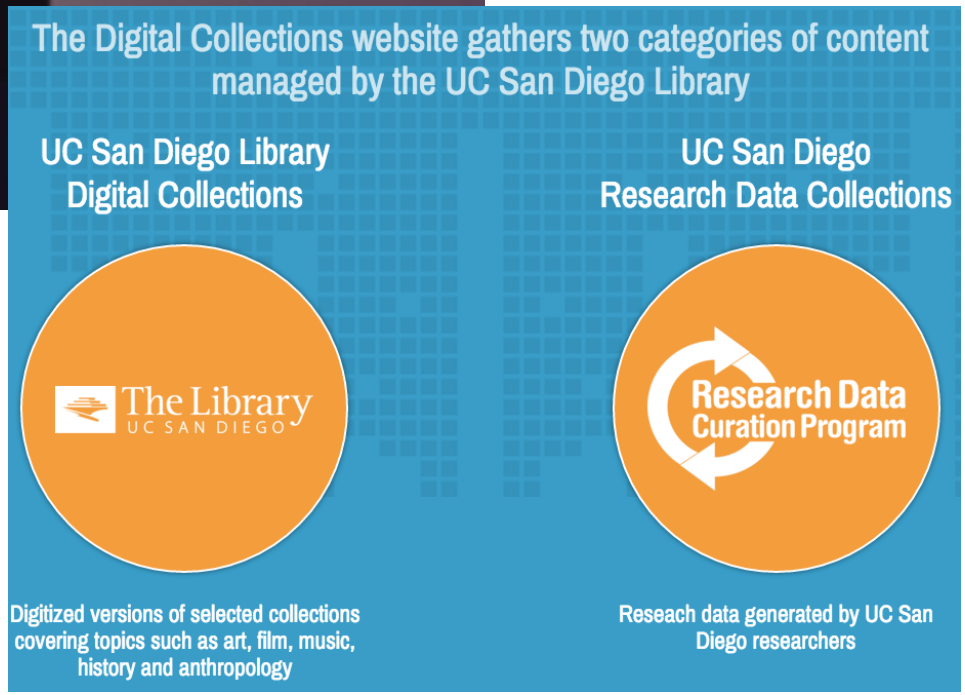
The Library Digital Collections

- Data is the intellectual capital of the modern research university. It belongs somewhere it can be kept safely and discoverable.

The Library
UC SAN DIEGO

The Library Digital Collections





http://library.ucsd.edu/dc

Browsing the Digital Collections

Browse by Collection: Library Digital Collections

View current collections below. Click an image or link for a full description.

Search Digital Collections

1 2 3 Go Sort: title 20 per page Advanced Search

[Baja California](#)

Images of Baja California digitized from original photographs held by UC San Diego Library. Includes landscapes, missions, the built environment, and ranch life.
1938-1997

Bear Facts

[Bear Facts](#)

Bear Facts was the newsletter published by the Oceanids, a service and social organization promoting friendship and services to the UC San Diego community.
1962-2007

[California Explores the Ocean](#)

A collection of diverse resources about the ocean and ocean exploration.

[Camp Matthews Photographs and Plans](#)

Photographs and plans of Camp Matthews buildings after Camp Matthews closed operations.
1964-1967

[Charles Atwood Kofoid Papers](#)

Photographs taken and collected by the University of California, Berkeley zoologist Charles Atwood Kofoid and his wife, Carrie Winter Kofoid. Subjects include the Kofoids; their families, friends and colleagues; travel; the 1904-1905 Albatross Expedition; and many freshwater and marine biological research stations.
circa 1860-1936

[Charles H. Graves Photographs and Films](#)

Film and color slides created for and used in a photographic documentary of early UCSD events.
1963-1967

[Chinese Cultural Revolution Posters](#)

Propaganda posters created during the Cultural Revolution by various Chinese agencies.
1966-1989

http://library.ucsd.edu/dc/rdcp/collections

http://library.ucsd.edu/dc/rdcp/collections

Browse by Collection: UCSD Research Data Collections

View current collections below. Click an image or link for a full description.

Search Digital Collections

Results 1 - 12 of 12 Sort: title 20 per page Advanced Search

[An annotated checklist of the bees (Hymenoptera: Anthophila) of San Diego County, California](#)

This database represents an ongoing survey to document all the bee species that occur within the bounds of San Diego County. As this project is still ongoing, the database will be updated periodically with additions and/or corrections.
July 2015,

[Bee Research Methods: Video Demonstrations](#)

This collection shows bee behaviors and methods used to study bee behaviors.
2005-2014

[California Tobacco Survey (CTS)](#)

The objective of these surveys is to collect representative statewide data on cigarette smoking behavior, attitudes towards smoking, media exposure to smoking, and use of tobacco products other than cigarettes, from populations living in California.
1990-

[Data from: Allostery through the Computational Microscope: cAMP Activation of a Canonical Signalling Domain](#)

This collection contains the molecular dynamics trajectories scripts and models used in the paper "Allostery through the computational microscope: cAMP activation of a canonical signalling domain".
2014-08-24,

[Data from: Carbonic Anhydrases, EPF2 and a Novel Protease Mediate CO2 Control of Stomatal Development](#)

RNA-Seq experiments analyzing the Arabidopsis response to low and elevated carbon dioxide growing conditions were conducted to capture early signaling events as seedlings were adapting to stress.
2012-2013

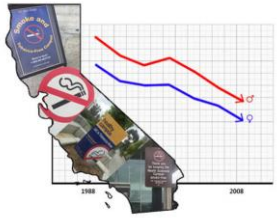
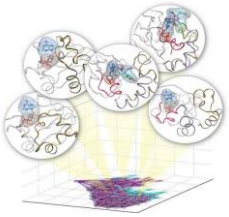
[Data from: Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning](#)

The kinetic rate constants of four biochemically relevant ligand-receptor systems were estimated and described in a publication. This collection includes input files, output files, trajectories, and statistics associated with the study.
2013-2015

Research Data Collections Stats

The Library UC SAN DIEGO

Collection	Items Count	Size (MB)
"Khirbat en-Nahas Project"@ar-Latn	1366	207,540
An annotated checklist of the bees (Hymenoptera: Anthophila) of San Diego County, California	1	0
Bee Research Methods: Video Demonstrations	5	1,068
California Tobacco Survey (CTS)	9	4,137
Data from: Allostery through the Computational Microscope: cAMP Activation of a Canonical Signalling Domain	1	19,686
Data from: Carbonic Anhydrases, EPF2 and a Novel Protease Mediate CO2 Control of Stomatal Development	1	31,005
Data from: Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning	1	268,703
Heavy Metals in the Ocean Insect, Halobates	2	92
Keith Rayner Eye Movements in Reading Data Collection	15	4.351
OpenTopography	103	4,810,330
SWell Ex-96 Experiment Acoustic Data	4	3,258
Santa Fe Light Cone Simulation research project files	39	751,579
Scripps Institution of Oceanography, Geological Collections	6761	4,798
Teaching Bee	12	259
TOTALS:	8,305	6,102,455


The Bees Got Us Started

The Library UC SAN DIEGO

DIGITAL COLLECTIONS

UC San Diego

Home | About | Browse By | Help | Sign In | The Library

Search Digital Collections

Collections »

Teaching Bee

About this collection

Description

This collection of documents provides teaching exercises and information for instructors of students ranging from junior high school through college. The target audience is indicated for each collection item.

Creator

James Nieh

Creation Date

2005-2013

Extent

12 digital objects.

Formats

View formats within this collection

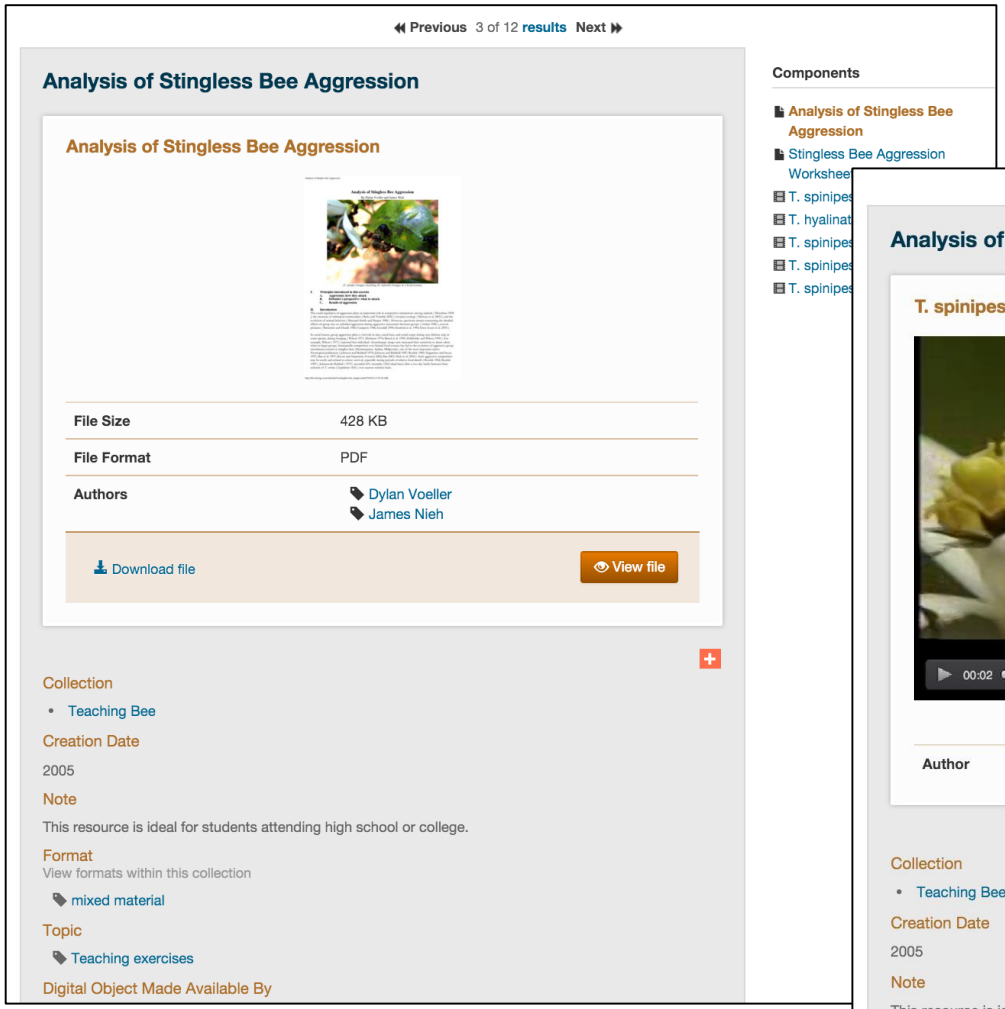
video
text

View Collection Items

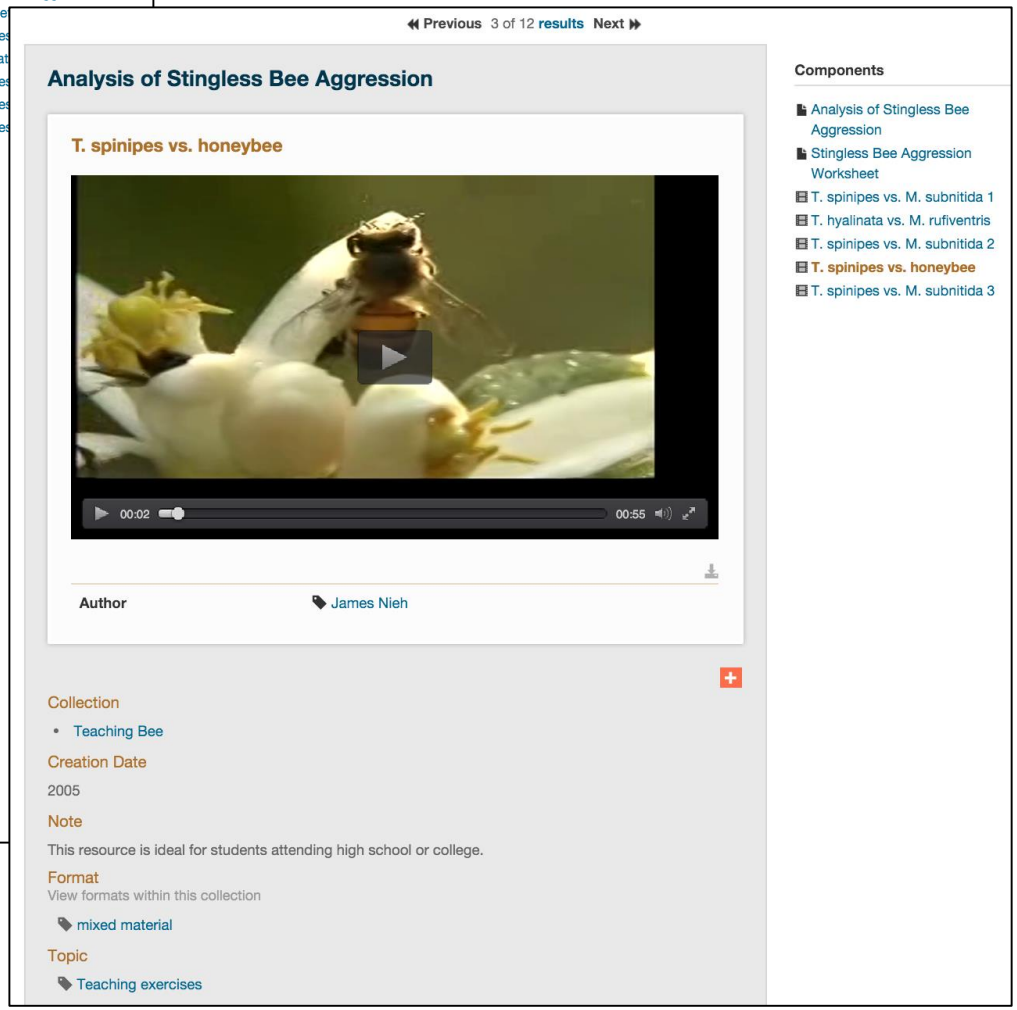
Search this collection...

http://library.ucsd.edu/dc/collection/bb6725113c

The Bees Got Us Started



Multimedia capable.



Downloadable files.

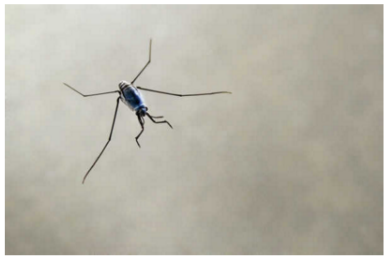
Unique Dataset

The Library UC San Diego

Collections »

Heavy Metals in the Ocean Insect, Halobates

About this collection



View Collection Items

Search this collection...

Description

Halobates is the only insect genus with representatives living their entire lives in the open ocean. It is a member of the true bug order Heteroptera and belongs to the family Gerridae which includes pond skaters commonly found in freshwater ponds, lakes, streams and rivers. Although the genus was first discovered in 1822, not much was known about its biology or special adaptations which enabled it to live in the open ocean where no other insects were able to survive. This was largely because few entomologists have any reasons to venture out to sea. Being attached to the Scripps Institution of Oceanography (SIO) with sea-going vessels has given me opportunities to study them at sea. What we now know about their distributions, biology, special adaptations and phylogeny can be found in reviews listed under 'Publication'.

This project was initiated when interests on heavy metal pollution Worldwide were generated following discoveries on lead poisoning from additives in petrol in the 1970s. Air pollutants eventually rain down to earth. Since the ocean covers more than 70% of the earth's surface, many chemical oceanographers were interested in finding out the presence and concentrations of heavy metals in the sea. It is relatively straightforward to measure heavy metal concentrations in seawater but measuring those that occur at the surface film which is the entry point of pollutants to the ocean is quite another matter. Methods and precautions needed in order to prevent contamination of surface samples were difficult to surmount. Halobates are completely wingless. They do not fly, nor dive throughout their lives. They are completely restricted to the sea-air interface. They were easy to collect and samples could be cleaned prior to measurements to give dependable results. They can therefore be useful indicators of heavy metal concentrations on the ocean film.

Principal Investigator

Cheng, Lanna

Co Principal Investigator

Schulz-Baldes, Meinhard

Contributors

Alexander, George V.
Franco, Paul J.
Ott, John

Date Issued

2015

Date Collected

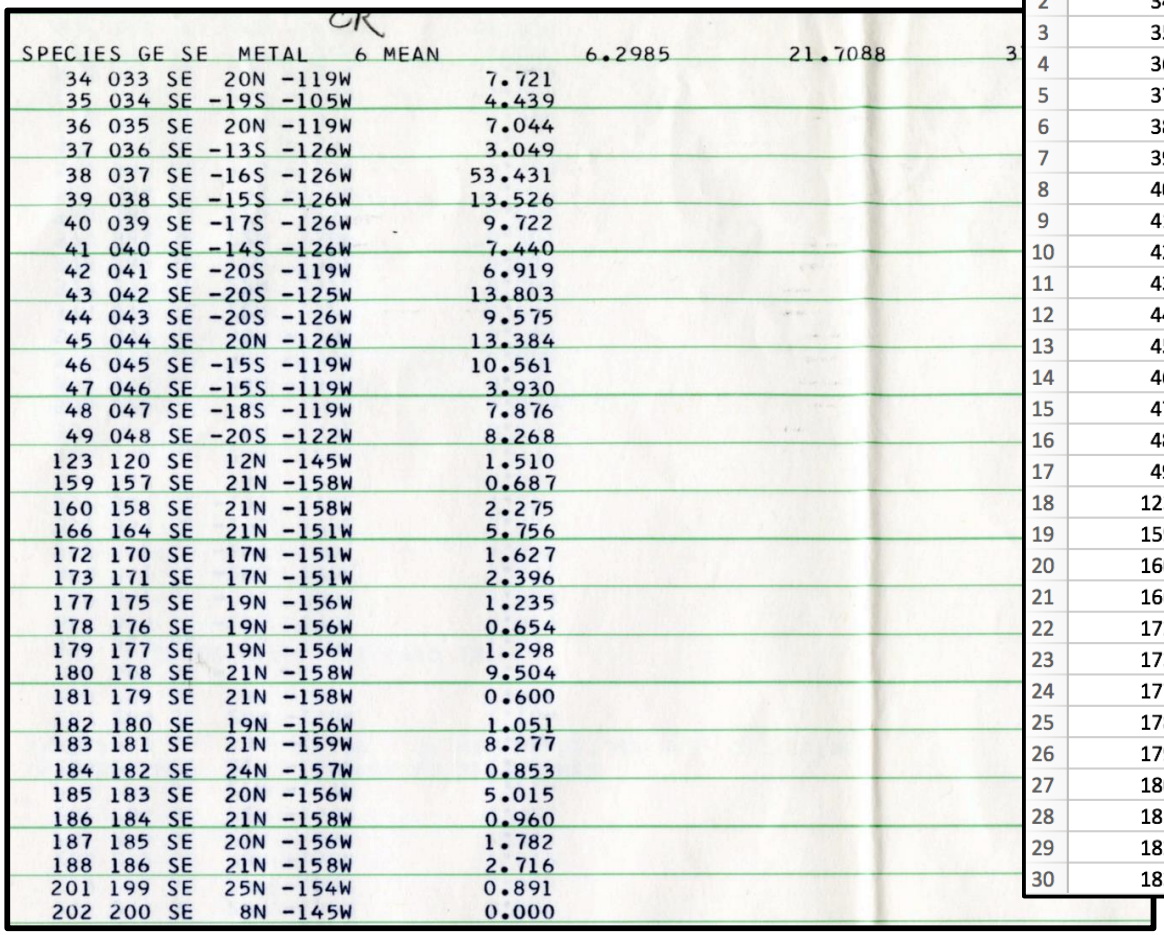
1961-1978

http://library.ucsd.edu/dc/collection/bb8056206n

Unique Dataset

Heavy Metals in the Ocean Insect, Halobates

- Scanned 63 computer printouts from the early 70's.
- OCR completed and verified.



CR

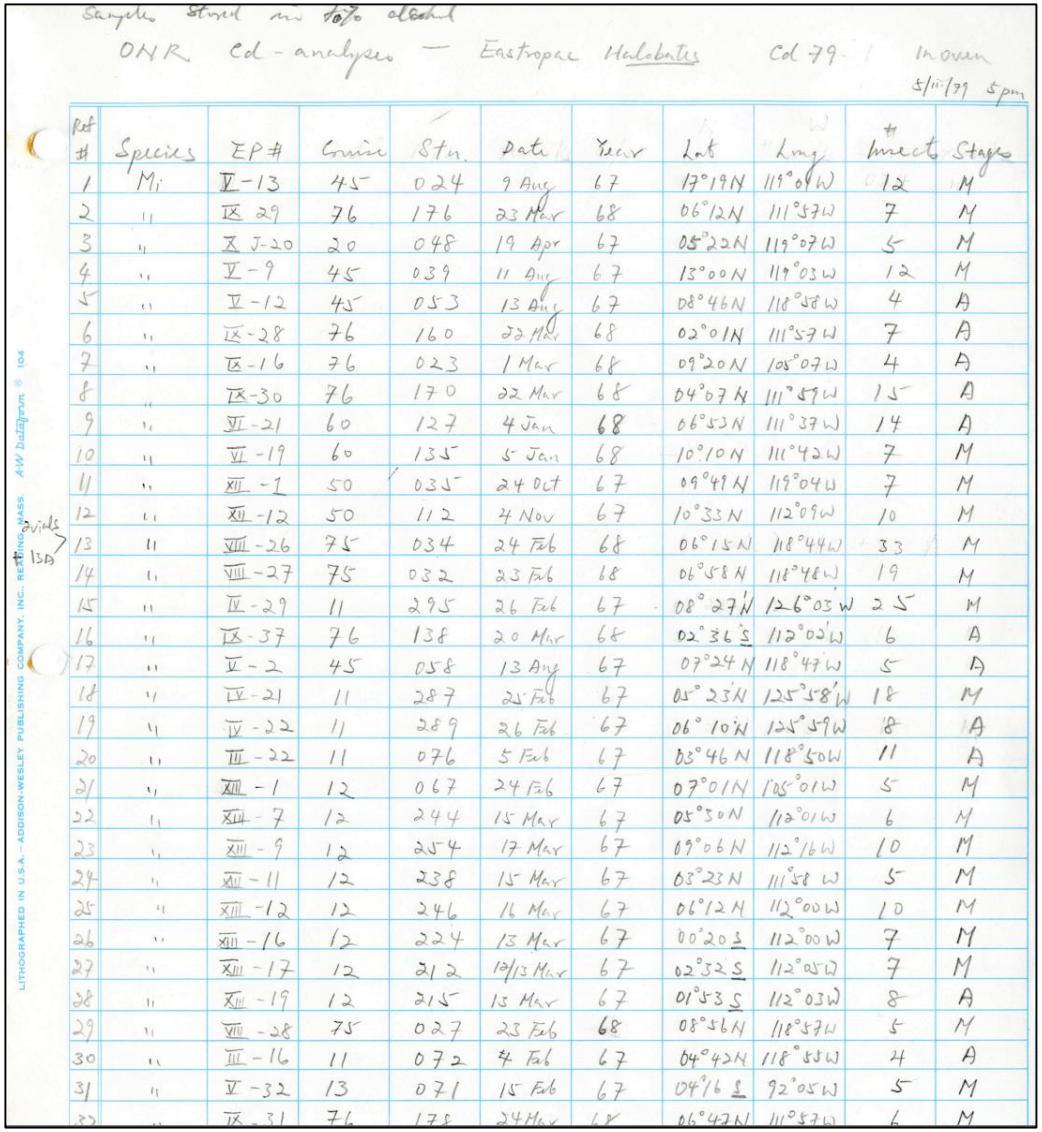
SPECIES	GE	SE	METAL	6 MEAN	6.2985	21.7088	3
34	033	SE	20N -119W	7.721			
35	034	SE	-19S -105W	4.439			
36	035	SE	20N -119W	7.044			
37	036	SE	-13S -126W	3.049			
38	037	SE	-16S -126W	53.431			
39	038	SE	-15S -126W	13.526			
40	039	SE	-17S -126W	9.722			
41	040	SE	-14S -126W	7.440			
42	041	SE	-20S -119W	6.919			
43	042	SE	-20S -125W	13.803			
44	043	SE	-20S -126W	9.575			
45	044	SE	20N -126W	13.384			
46	045	SE	-15S -119W	10.561			
47	046	SE	-15S -119W	3.930			
48	047	SE	-18S -119W	7.876			
49	048	SE	-20S -122W	8.268			
123	120	SE	12N -145W	1.510			
159	157	SE	21N -158W	0.687			
160	158	SE	21N -158W	2.275			
166	164	SE	21N -151W	5.756			
172	170	SE	17N -151W	1.627			
173	171	SE	17N -151W	2.396			
177	175	SE	19N -156W	1.235			
178	176	SE	19N -156W	0.654			
179	177	SE	19N -156W	1.298			
180	178	SE	21N -158W	9.504			
181	179	SE	21N -158W	0.600			
182	180	SE	19N -156W	1.051			
183	181	SE	21N -159W	8.277			
184	182	SE	24N -157W	0.853			
185	183	SE	20N -156W	5.015			
186	184	SE	21N -158W	0.960			
187	185	SE	20N -156W	1.782			
188	186	SE	21N -158W	2.716			
201	199	SE	25N -154W	0.891			
202	200	SE	8N -145W	0.000			

- Manually verified and tested Excel data entries

	A	B	C	D	E	F	G
1	sequence	specimen_I	species	latitude	longitude	mean_conc	metal
2	34	033	SE	20N	-119W	7.721	Cr
3	35	034	SE	-19S	-105W	4.439	Cr
4	36	035	SE	20N	-119W	7.044	Cr
5	37	036	SE	-13S	-126W	3.049	Cr
6	38	037	SE	-16S	-126W	53.431	Cr
7	39	038	SE	-15S	-126W	13.526	Cr
8	40	039	SE	-17S	-126W	9.722	Cr
9	41	040	SE	-14S	-126W	7.440	Cr
10	42	041	SE	-20S	-119W	6.919	Cr
11	43	042	SE	-20S	-125W	13.803	Cr
12	44	043	SE	-20S	-126W	9.575	Cr
13	45	044	SE	20N	-126W	13.384	Cr
14	46	045	SE	-15S	-119W	10.561	Cr
15	47	046	SE	-15S	-119W	3.930	Cr
16	48	047	SE	-18S	-119W	7.876	Cr
17	49	048	SE	-20S	-122W	8.268	Cr
18	123	120	SE	12N	-145W	1.510	Cr
19	159	157	SE	21N	-158W	0.687	Cr
20	160	158	SE	21N	-158W	2.275	Cr
21	166	164	SE	21N	-151W	5.756	Cr
22	172	170	SE	17N	-151W	1.627	Cr
23	173	171	SE	17N	-151W	2.396	Cr
24	177	175	SE	19N	-156W	1.235	Cr
25	178	176	SE	19N	-156W	0.654	Cr
26	179	177	SE	19N	-156W	1.298	Cr
27	180	178	SE	21N	-158W	9.504	Cr
28	181	179	SE	21N	-158W	0.600	Cr
29	182	180	SE	19N	-156W	1.051	Cr
30	183	181	SE	21N	-159W	8.277	Cr

Unique Dataset

Heavy Metals in the Ocean Insect, Halobates



Samples stored in 70% alcohol

ONR Cd - analyses — Eastropac Halobates Cd 79 In oven 5/11/79 5pm

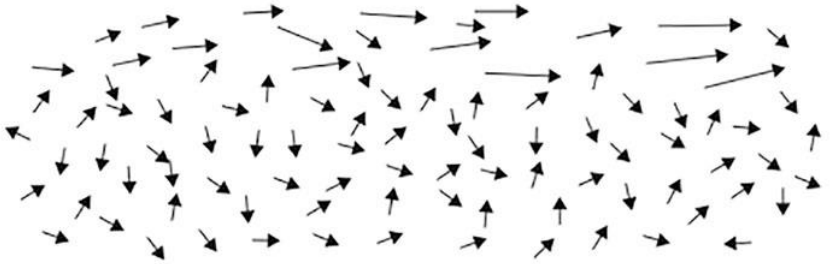
Ref #	Species	EP #	Cruise	Stn.	Date	Year	Lat	Long	# Insect	Stages
1	Mi	V-13	45	024	9 Aug	67	17°19N	119°09W	12	M
2	"	IX 29	76	176	23 Mar	68	06°12N	111°57W	7	M
3	"	X J-20	20	048	19 Apr	67	05°22N	119°07W	5	M
4	"	V-9	45	039	11 Aug	67	13°00N	119°03W	12	M
5	"	V-12	45	053	13 Aug	67	08°46N	118°58W	4	A
6	"	IX-28	76	160	22 Mar	68	02°01N	111°57W	7	A
7	"	IX-16	76	023	1 Mar	68	09°20N	105°07W	4	A
8	"	IX-30	76	170	22 Mar	68	04°07N	111°59W	15	A
9	"	VI-21	60	127	4 Jan	68	06°53N	111°37W	14	A
10	"	VI-19	60	135	5 Jan	68	10°10N	111°42W	7	M
11	"	XII-1	50	035	24 Oct	67	09°49N	119°04W	7	M
12	"	XII-12	50	112	4 Nov	67	10°33N	112°09W	10	M
13	"	VIII-26	75	034	24 Feb	68	06°15N	118°44W	33	M
14	"	VIII-27	75	032	23 Feb	68	06°58N	118°48W	19	M
15	"	IV-29	11	295	26 Feb	67	08°27'N	126°03'W	25	M
16	"	IX-37	76	138	20 Mar	68	02°36'S	112°02'W	6	A
17	"	V-2	45	058	13 Aug	67	07°24N	118°47W	5	A
18	"	IV-21	11	287	25 Feb	67	05°23'N	125°58'W	18	M
19	"	IV-22	11	289	26 Feb	67	06°10'N	125°59W	8	A
20	"	III-22	11	076	5 Feb	67	03°46N	118°50W	11	A
21	"	XIII-1	12	067	24 Feb	67	07°01N	105°01W	5	M
22	"	XIII-7	12	244	15 Mar	67	05°30N	112°01W	6	M
23	"	XIII-9	12	254	17 Mar	67	09°06N	112°16W	10	M
24	"	XIII-11	12	238	15 Mar	67	08°23N	111°58W	5	M
25	"	XIII-12	12	246	16 Mar	67	06°12N	112°00W	10	M
26	"	XIII-16	12	224	13 Mar	67	00°20'S	112°00W	7	M
27	"	XIII-17	12	212	12/13 Mar	67	02°32'S	112°05W	7	M
28	"	XIII-19	12	215	13 Mar	67	01°53'S	112°03W	8	A
29	"	VIII-28	75	027	23 Feb	68	08°56N	118°57W	5	M
30	"	III-16	11	072	4 Feb	67	04°42N	118°55W	4	A
31	"	V-32	13	071	15 Feb	67	04°16'S	92°05W	5	M
32	"	IX-31	76	178	24 Mar	68	06°47N	111°57W	6	M

LITHOGRAPHED IN U.S.A. - ADDISON-WESLEY PUBLISHING COMPANY, INC. - READING, MASS. A-W Dataform ® 104

Hand written collection event record was added as a part of the dataset.

Buzzing Ahead

- Improve existing services and Increase the number of research data curated by the library.
- Further streamlining the data collection workflow.
- Continue pursuing partnerships with academic departments and researchers through outreach with our liaison librarians.
- Implement new training ideas and workshops to encourage more use of our tools and services.



Side view

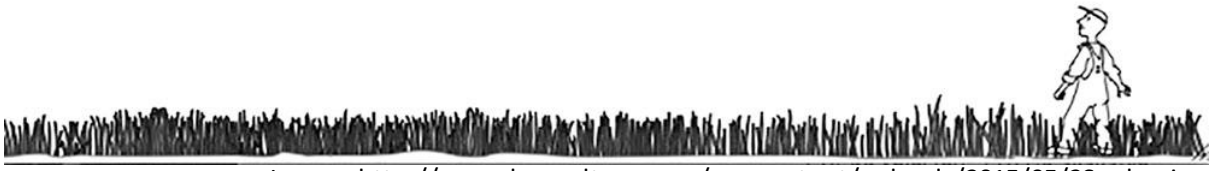


Image: http://www.beeculture.com/wp-content/uploads/2015/05/2Seeley.jpg

Thank you!

The Library UC SAN DIEGO

Questions?

Reid Otsuji
Data Curation Specialist and Faculty Liaison
UC San Diego Library Research Data Curation Program
rotsuji@ucsd.edu

UC San Diego Library Digital Collections
http://library.ucsd.edu/dc

The Library Research Data Curation Program, UC San Diego
http://lib.ucsd.edu/rdcp